

Exhibit C

Kenney, Chad (DC)

From: Bennett, Allyson R. <ABennett@mofo.com>
Sent: Wednesday, May 15, 2024 11:12 PM
To: Zach Savage; NYT-AI-SG-Service@simplelists.susmangodfrey.com; Ranieri, Vera; Alex Frawley; Nikolaides, Eric K.; OpenAICopyright; #C-M OPENAI COPYRIGHT LITIGATION - LW TEAM; NewYorkTimes_Microsoft_OHS@orrick.com; MicrosoftNYClassActionFDBR@faegredrinker.com; KVPOAI@keker.com
Subject: RE: The New York Times Company v. Microsoft Corporation, et al. - SDNY Case No 1:23-cv-11195-SHS

Zach,

Please see below for OpenAI's responses to the points raised in your May 8 email.

Search Terms and Custodians

OpenAI's investigation is ongoing. That said, OpenAI is willing to pick a date on which all parties will mutually exchange their custodians and search terms. Please let me know when you will be prepared to complete the exchange.

Document Production

OpenAI is working diligently to review documents and prepare them for production. OpenAI is prepared to make its first production in short order, subject to the following: In your email you stated that (notwithstanding that the Court has not yet entered either party's protective order), Plaintiff would treat any documents produced by OpenAI as "confidential or highly confidential, as the case may be, under the parties' draft protective orders." Please confirm that if we produce documents, you will afford them the protection provided under *OpenAI's* proposed protective order, not Plaintiff's. Further, we note that Plaintiff also has not produced any documents yet. When can we expect Plaintiff's first production? Relatedly, please also let us know by Thursday your final position regarding whether you will agree to a June 24 substantial completion date for the NYT's production.

Limitation to Inspection of "Pretraining Data" (RFP Nos. 1, 3, 6, 8, 14)

Making available for inspection all data used for "post-training" —a term we use here to refer to model training occurring after the pre-training phase, like fine-tuning— would be unduly burdensome. OpenAI is, however, conducting a reasonable search for any New York Times data used for post-training, and will make any such data found as a result of that search available for inspection. Thus far, OpenAI's search has not located any New York Times information used for post-training, but we will let you know if that changes.

Moreover, information about models not used for ChatGPT is irrelevant and, in any event, it would be overly burdensome to collect such information. As stated publicly by OpenAI, each new generation of foundation models is trained from scratch. See OpenAI, Our approach to data and AI, <https://openai.com/index/approach-to-data-and-ai> (May 7, 2024).

As to RFP 14, we continue to believe that permitting inspection of the relevant pretraining data for models used for ChatGPT is sufficient. After inspecting the relevant data, if Plaintiff believes OpenAI's response is deficient, it can identify what information it believes is missing and OpenAI will respond appropriately at that time.

Subject to the above disputes and as stated in our May 4 letter, OpenAI will consider using custodians and search terms to conduct a reasonable search for documents responsive to RFPs 1 and 3, including any non-custodial documents that exist as well as custodial documents, that are nonprivileged and can be located after a reasonable search.

Limiting Definitions of Generative AI Products and Services and AI Models (RFP Nos. 5, 12, 13)

Browse (fka Browse with Bing) is an OpenAI product that is a feature within ChatGPT, and OpenAI's proposed definition of "Generative AI Products and Services" and "AI Models" therefore encompasses Browse. Other products, such as Bing Chat and Copilot, are third-party products, and the Times has not demonstrated how these products are relevant to its claims against OpenAI or why it would not be overly burdensome for OpenAI to search for that information. If the Times can demonstrate relevance, OpenAI is open to discussing relevant search terms and custodians.

Policies, Procedures & Practices regarding Intellectual Property in AI Models (RFP No. 5)

As discussed in the meet and confer, the bulk of documents relevant to this request are likely privileged. Setting aside privileged documents, we are still unsure what documents you are looking for other than Copyright Shield. Based on our discussion, we understand you are also interested in opt-out or filtering-related documents. OpenAI is willing to search for and produce documents sufficient to show OpenAI's filtering of pre-training data and its policies with respect to opting out.

Narrowing of RAG-Related Requests (RFP Nos. 12, 13)

As discussed above, OpenAI maintains that requests for documents relating to third-party products are not relevant, and OpenAI will not search for and produce documents regarding models not developed by OpenAI. Because Browse is an OpenAI product, OpenAI will accordingly produce RAG-related documents, if any, pertaining to Browse. Moreover, our response agrees to produce documents "sufficient to show how OpenAI uses RAG in connection with ChatGPT." Therefore, to the extent relevant documents exist beyond Browse, they would fit within our response.

OpenAI confirms that it will produce documents "sufficient to show how OpenAI uses RAG in connection with ChatGPT." As you know from our discussion, the "sufficient to show" limitation is necessary because RAG is a broad and general term that would sweep in a lot of irrelevant documents. Such a search would be overbroad, unduly burdensome, and disproportionate to the needs of the case. As noted on the meet and confer, however, if Plaintiff has anything more specific in mind with respect to this RFP (in addition to Browse), please let us know and we will consider it.

Additionally, while Plaintiff's email does not address the issue, OpenAI is willing to produce non-privileged documents sufficient to show whether OpenAI excludes sources known to have paywalls.

Legislative or Executive Inquiries (RFP No. 2)

There is no legislative or executive inquiry whose primary focus is copyright issues. Accordingly, there is no "list" of relevant inquiries for OpenAI to produce. In any event, producing such a list would be overly burdensome, and is not proportional to the needs of the case.

Claims of Copyright Infringement (RFP No. 4)

As we stated in our original RFP responses and subsequent letter, OpenAI limited its response to "formal, written complaints," if any, because it was unclear what else Plaintiff wanted, or how OpenAI could feasibly search for anything else. Plaintiff's view that "OpenAI should also search for and produce communications addressing complaints or allegations of copyright and trademark infringement, whether 'formal' or informal, including internal OpenAI documents and communications as well as complaints raised to OpenAI by external entities," is overbroad, unduly burdensome, and disproportionate to the needs of the case. Moreover, it likely seeks documents that are protected by privilege and/or are irrelevant, and Plaintiff refuses to produce similar types of documents that are more relevant to the claims and

defenses in this case (*see, e.g.*, OpenAI RFPs to NYT 8, 9, 10). If Plaintiff contends OpenAI's response is deficient, please explain why the Times's response is not also deficient.

Subject to these objections, OpenAI will draft and run search terms for documents constituting (1) complaints of copyright infringement, if any, regarding the use of purportedly copyrighted material or (2) OpenAI's responses, if any, to such complaints.

Alternatives to Using Copyrighted Content (RFP No. 6)

OpenAI's data access agreements do not differentiate between types of training data. Thus, the agreements OpenAI has already agreed to produce are also related to fine-tuning. Regardless, OpenAI reserves all rights concerning the relevance of fine-tuning datasets.

As we mentioned at the meet and confer, many communications regarding agreements are likely protected by privilege. Further, searching for such documents would be disproportionate to the needs of the case, particularly given that the Times has itself refused to search for similar documents.

Impact of AI Models on Journalism (RFP No. 7)

OpenAI has agreed to produce both custodial and non-custodial documents for this request, subject to the parties' disagreement on the definition of "Generative AI Products and Services and AI Models," after the entry of a protective order in this case.

OpenAI's Transition to a For-Profit Company (RFP No. 11)

OpenAI understands that Plaintiffs have abandoned their position that any OpenAI entity has "transitioned" to a for-profit company, given no such transition has occurred. And even if such a transition had occurred, documents regarding the transition itself would not be relevant, and searching for documents related to that transition would be overly burdensome. Nevertheless, we will produce documents sufficient to show the corporate formation of relevant OpenAI entities and their relationship to each other.

Collaborations Between Defendants Relating to the Generative AI Models (RFP No. 15)

OpenAI understands that Microsoft has already agreed to produce "responsive, nonprivileged documents concerning any commercial or collaborative efforts or contractual agreements between Microsoft and OpenAI relating to OpenAI's Text Generation AI Models identified in the Complaint." Please identify what documents, if any, you seek and that you are not already receiving from Microsoft.

Thanks,

Ally

ALLYSON BENNETT (she/her)

Partner

abennett@mofo.com

T +1 (213) 892-5200

M +1 (818) 294-3559

**MORRISON
FOERSTER**